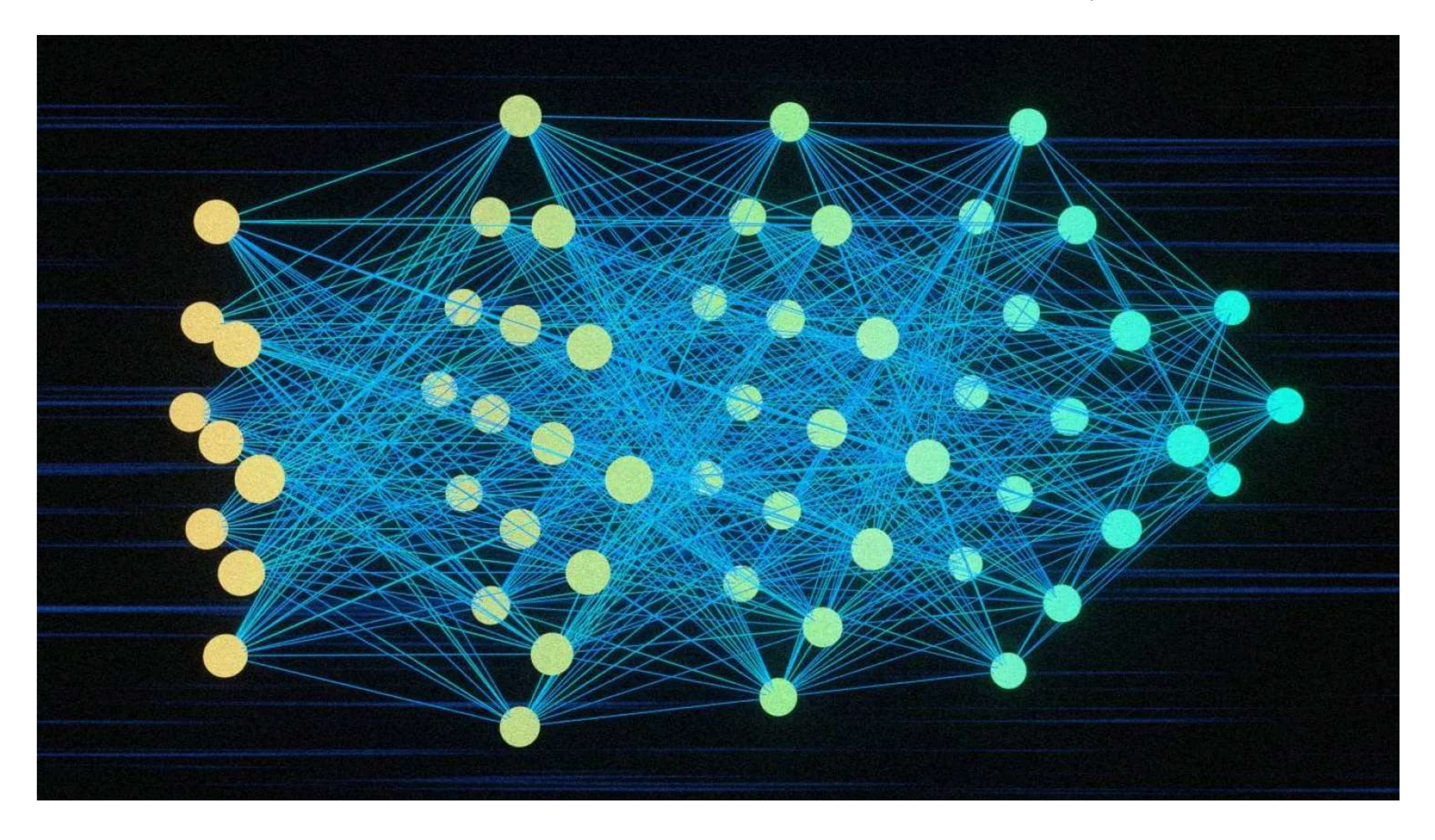
علوم

باحثون أميركيون يسبرون أغوار «تفكير» الذكاء الاصطناعي التوليدي

نشاط غامض ناجم عن إشارات رياضية وشبكات عصبية



واشنطن: مارك سوليفان

نُشر: 12:26-2 أبريل 2025 م . 04 شوّال 1446 هـ

لماذا تُعد روبوتات الدردشة القائمة على الذكاء الاصطناعي ذكية للغاية، فهي قادرة على فهم الأفكار المعقدة، وصياغة قصص قصيرة رائعة بشكل مدهش، وفهم ما يقصده المستخدمون بديهياً؟ الحقيقة هي أننا لا نعرف الإجابة تماماً.

نشاط فكري غامض

تُفكر نماذج اللغة الكبيرة بطرق لا تبدو بشرية تماماً، إذ تتكون مخرجاتها من مليارات الإشارات الرياضية التي تتدفق عبر طبقات من الشبكات العصبية التي تعمل بأجهزة كمبيوتر ذات قوة وسرعة غير مسبوقتين. ويظل معظم هذا النشاط غير مرئي أو غامض بالنسبة لباحثي الذكاء الاصطناعي.

يُمثل هذا الغموض تحديات واضحة، لأن أفضل طريقة للتحكم في شيء ما، فهم كيفية عمله.

لقد كان لدى العلماء فهم راسخ للفيزياء النووية قبل بناء أول قنبلة أو محطة طاقة. ولكن لا يمكن قول الشيء نفسه عن نماذج الذكاء الاصطناعي التوليدية. ولا يزال الباحثون العاملون في مجال سلامة الذكاء الاصطناعي، وهو فرع من فروع «التفسير الآلي»، والذين يقضون أيامهم في دراسة التسلسلات المعقدة للدوال الرياضية التي تُؤدي إلى إخراج هذه النماذج للكلمة أو البكسل (عنصر الصورة) التالي، يحاولون اللحاق بالركب.

أبحاث جديدة

الخبر السار هو أنهم يُحرزون تقدماً حقيقياً. ومثال على ذلك: إصدار بحثين جديدين من شركة أنثروبيك، يُقدمان رؤى جديدة حول «التفكير» الداخلي لبرنامج الدردشة الذكي.

«مجهر» لتدقيق الذكاء الاصطناعي

وكما تعتمد المعاملات داخل الشبكات العصبية على «الخلايا العصبية» في الدماغ، استعان باحثو أنثروبيك بعلم الأعصاب لدراسة الذكاء الاصطناعي. وصرّح جوشوا باتسون، عالم الأبحاث في «أنثروبيك»، لمجلة «فاست كومباني»، بأن فريقه طوّر أداة بحثية - أشبه بـ«مجهر الذكاء الاصطناعي» - يمكنها تتبع أنماط البيانات وتدفقات المعلومات داخل البرنامج الذكي، ومراقبة كيفية ربطه للكلمات والمفاهيم في طريقه إلى الإجابة.

قبل عام، لم يكن بإمكان الباحثين سوى رؤية سمات محددة لهذه الأنماط والتدفقات، لكنهم بدأوا الآن في ملاحظة كيف تُؤدي فكرة إلى أخرى من خلال سلسلة من التفكير المنطقي. ويقول باتسون: «نحاول ربط كل ذلك معاً، ونشرح خطوة بخطوة عند وضع مُوجِّه في نموذج، لماذا يقول الكلمة التالية؟ وبما أن إجابات النموذج تأتي كلمة تلو أخرى، فإذا استطعت تحليلها والقول: حسناً، لماذا قال هذه الكلمة بدلاً من تلك؟ يمكنك حينها فهم الأمر برمته».

الذكاء الاصطناعي والرياضيات

الذكاء الاصطناعي يُفكِّر بشكل مختلف - حتى عندما يتعلق الأمر بالرياضيات البسيطة.

ويُعززِ هذا البحث فكرة أن أنظمة الذكاء الاصطناعي تُعالج المشكلات بشكل مختلف تماماً عن البشر. لا تُدرس الأنظمة مهام مثل الحساب بشكل صريح، بل تُعرَض عليها الإجابات الصحيحة، وتترك لتطوير مسارها الاحتمالي الخاص نحو هذا الاستنتاج.

درس باتسون وفريقه مثالاً بسيطاً على هذه الرياضيات - طلب من برنامج دردشة جمع العددين 36 و59 - ووجد أن «عملية» الذكاء الاصطناعي كانت مختلفة تماماً عن حسابات الإنسان العادي. فبدلاً من اتباع خطواتٍ بشرية، استخدم نموذج الاختبار نوعين من المنطق للوصول إلى الإجابة: تقريب الإجابة (هل هي في التسعينات؟) وتقدير الرقم الأخير منها. بجمع احتمالات الإجابات المختلفة، تمكن برنامج «كلود» الذي طورته شركة «أنثروبيك» من الوصول إلى المجموع الصحيح. يقول باتسون: «لقد تعلم بالتأكيد استراتيجية مختلفة في الرياضيات عن تلك التي تعلمناها في المدرسة».

التفكير بمفاهيم شمولية

درس الباحثون أيضاً ما إذا كانت برامج التعلم الآلي، التي غالباً ما تُحلل وتُنتج محتوى بلغات متعددة، تُفكّر بالضرورة بلغة الكلمات المُعطاة لها التي يوجهها المستخدم. يتساءل باتسون: «هل تستخدم الكلمات الإنجليزية فقط عند التعامل مع اللغة الإنجليزية، والأجزاء الفرنسية عند التعامل مع اللغة الفرنسية، والأجزاء الصينية عند التعامل مع اللغة الصينية؟... أم أن هناك أجزاءً من النموذج تُفكّر بالفعل بمفاهيم شمولية عالمية بغض النظر عن اللغة التي تعمل بها؟».

رموز عالمية تترجم إلى اللغات

وجد الباحثون أن برامج التعلم الآلي تقوم بكلا الأمرين. طلبوا من «كلود» ترجمة جمل بسيطة إلى لغات متعددة، وتتبعوا الرموز المتداخلة التي استخدمها أثناء المعالجة. تُمثّل هذه الرموز المُشتركة - أي مقتطفات من المعنى - أفكاراً جوهرية لا تعتمد على لغة مُحددة، مثل «الصغر» أو «التضاد». وقد أدى استخدام هذين الرمزين معاً إلى تمثيل مفهوم عالمي آخر يُمثّل «الكبر» (عكس الصغير هو الكبير). يستخدم النموذج هذه المفاهيم العالمية قبل أن يترجمها إلى لغة معينة للمستخدم.

يشير هذا إلى أن «كلود» يستطيع تعلم مفهوم مثل «الصغر» في لغة ما، ثم تطبيق هذه المعرفة عند التحدث بلغة أخرى دون أي تدريب إضافي، كما يقول باتسون. تُعدّ دراسة كيفية مشاركة النموذج لما يعرفه عبر السياقات أمراً مهماً لفهم طريقة تفكيره في الأسئلة في العديد من المجالات المختلفة.

نظم ذكية في التخطيط والارتجال

لا يفكر «كلود» فقط في الكلمة المنطقية التالية التي يجب توليدها، بل لديه أيضاً القدرة على التفكير «مستقبلاً». عندما طلب منه فريق البحث كتابة الشعر، أدرج «كلود» بالفعل أنماط القافية في أنماط معالجته. على سبيل المثال، بعد أن انتهى السطر بالعبارة الإنجليزية «grab it»، اختار «كلود» كلمات في السطر التالي من شأنها أن تُهيئ بشكل جيد لاستخدام كلمة «rabbit» خاتمة.

الذكاء الاصطناعي ونَظْم القصائد الشعرية

يقول باتسون: «وجد أحد أعضاء فريقي أنه في نهاية هذا السطر، بعد (grab it)، وقبل أن يبدأ حتى في كتابة السطر التالي، كان يفكر في أرنب (rabbit)»، تدخل الباحثون بعد ذلك في تلك المرحلة تحديداً من العملية، فأدخلوا إما نظام قافية جديداً أو كلمة ختامية جديدة، وقام (كلود) بتغيير خطته وفقاً لذلك، مختاراً مساراً لفظياً جديداً للوصول إلى قافية منطقية.

يقول باتسون إن ملاحظة الشعر هي من مفضلاته لأنها تعطي نظرة واضحة نسبياً على جزء محدد من تفكير البرنامج الذكي في حل مشكلة ما، ولأنها تثبت أن أدوات الملاحظة التي استخدمها فريقه (مثل مجهر الذكاء الاصطناعي) تؤدي عملها.

تسلط دراسة الشعر الضوء على مقدار العمل الذي لا يزال يتعين القيام به.

يلتقط الباحثون في هذا الميدان لقطات سريعة، بنفس الطريقة التي قد يدرس بها عالم أعصاب الطريقة التي يحول بها إحدى مناطق «الحُصين (hippocampus)» البشري الذكريات قصيرة المدى إلى ذكريات طويلة المدى. يقول باتسون: «استكشاف هذا المجال المعقد أشبه بمغامرة في كل مرة، ولذلك كنا في الواقع نحتاج فقط إلى أدوات لفهم كيفية ترابط الأشياء وتجربة الأفكار والتنقل بينها... لذا، نمر بمرحلة التحقيق هذه بعد بناء المجهر، وننظر إلى شيء ما ونقول: حسناً، ما هذا الجزء؟ وما هذا الجزء؟ وما هذا الشيء هنا؟».

التوجيه نحو سلوك آمن

ولكن بافتراض أن شركات الذكاء الاصطناعي تواصل تمويل أبحاث قابلية التفسير الآلي وإعطائها الأولوية، فإن اللقطات ستتوسع وتبدأ في الترابط، ما يوفر فهماً أوسع لسبب ما تفعله البرامج الذكية.

إن الفهم الأفضل لهذه الأنماط يمكن أن يمنح الباحثين فهماً أفضل للمخاطر الحقيقية التي قد تشكلها هذه الأنظمة، بالإضافة إلى طرق أفضل «لتوجيه» الأنظمة نحو سلوك آمن وخير.

يشير باتسون إلى أننا قد نطور ثقة أكبر بأنظمة الذكاء الاصطناعي بمرور الوقت من خلال اكتساب المزيد من الخبرة في مخرجاتها. ومع ذلك، يضيف أنه سيكون «أكثر ارتياحاً بكثير إذا فهمنا أيضاً ما يجري (في الداخل)».

* مجلة فاست كومباني، خدمات «تريبيون ميديا».

مواضيع

الذكاء الاصطناعي

العالم